Iran University of Science and Technology

Automotive Science and Engineering

Journal Homepage: ase.iust.ac.ir

ISSN: 2008-9899

Ethical Decision-Making in Autonomous Vehicles: A Human-Centric Risk Mitigation Approach Using Deep Q-Networks

Ehsan Vakili [1], Behrooz Mashadi [2], Abdollah Amirkhani [3*]

School of Automotive Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran.

ARTICLE INFO

Article history:
Received: 27 Oct 2024
Accepted: 15 Jan 2025
Published: 3 Feb 2025

Keywords:
Autonomous vehicles
Ethical decision-making
Human-centric
Risk mitigation
DQN

A B S T R A C T

Ensuring that ethically sound decisions are made under complex, real-world conditions is a central challenge in deploying autonomous vehicles (AVs). This paper introduces a human-centric risk mitigation framework using Deep Q-Networks (DQNs) and a specially designed reward function to minimize the likelihood of fatal injuries, passenger harm, and vehicle damage. The approach uses a comprehensive state representation that captures the AV's dynamics and its surroundings (including the identification of vulnerable road users), and it explicitly prioritizes human safety in the decision-making process. The proposed DQN policy is evaluated in the CARLA simulator across three ethically challenging scenarios: a malfunctioning traffic signal, a cyclist's sudden swerve, and a child running into the street. In these scenarios, the DQN-based policy consistently minimizes severe outcomes and prioritizes the protection of vulnerable road users, outperforming a conventional collision-avoidance strategy in terms of safety. These findings demonstrate the feasibility of deep reinforcement learning for ethically aligned decision-making in AVs and point toward a pathway for developing safer and more socially responsible autonomous transportation systems.

1. Introduction

A significant transformation of contemporary transportation systems is expected to be driven by autonomous vehicles (AVs), which offer enhanced safety, efficiency, and accessibility [1-3]. However, an exponential increase in the need to ensure ethically sound decision-making has been observed as these systems transition from controlled testing grounds to public roads [4]. Traditional rule-based or machine learning approaches often focus on optimizing technical objectives such as fuel efficiency or travel time, under the implicit assumption that standard collision-avoidance algorithms guarantee sufficient safety [5, 6]. Nevertheless, recent high-profile incidents and regulatory pressures have highlighted the importance of explicitly addressing ethical trade-offs, particularly in high-risk scenarios where human life and well-being are at stake [7].

Within the broader field of automated driving research, theoretical scenarios (e.g., variations of the "trolley problem") have largely guided investigations into AV ethics [8]. Although such scenarios capture moral complexity, few practical computational methods have been proposed to

*Corresponding Author
Email Address: amirkhani@iust.ac.ir
https://doi.org/10.22068/ase.2025.697

[DOI: 10.22068/ase.2025.697]

operationalize ethical principles in real-time control policies [9]. Reinforcement learning (RL) methods, especially DQNs, have been applied with some success to manage the complexity of dynamic driving environments [10]. However, a reliance on reward functions that emphasize technical metrics, such as maximizing driving comfort or minimizing average collision frequency, is frequently observed, and the balancing of potential harm to different human stakeholders is often neglected [11].

In this paper, a human-centric risk mitigation framework is proposed, in which DQNs are leveraged to support ethically informed decision-making in AVs. By incorporating a comprehensive state space—encompassing both the ego vehicle's kinematic information and the relative positions, velocities, and classifications of nearby objects—situational awareness is reinforced within the decision process. A carefully crafted reward function is also introduced to minimize fatal injury probabilities, passenger risk, and vehicular damage, thereby placing human safety at the forefront. To evaluate the effectiveness of the framework, ethically charged scenarios are constructed within the CARLA simulator [12], including malfunctioning traffic signals, sudden cyclist swerves, and the unpredictable entry of a child into the street. Through these evaluations, it is demonstrated that a DQN-based strategy can successfully navigate complex traffic environments while aligning decisions with human-centric ethical principles.

2. Related Work

Research on ethical decision-making in autonomous vehicles (AVs) has been increasingly highlighted due to the complexity involved in translating moral principles into computational models [8]. The literature has primarily focused on defining ethical frameworks, examining various forms of risk, and investigating artificial intelligence (AI) methods—particularly reinforcement learning (RL)—that can operationalize these frameworks in real-time driving contexts. In this section, key strands of the existing research are reviewed to showcase current challenges and proposed solutions within ethically guided AV decision-making.

In early investigations, moral dilemmas such as the "trolley problem"—which require choosing between multiple harmful outcomes—were used to illustrate the ethical conundrums faced by autonomous vehicles (AVs). The trolley problem remains a central thought experiment in evaluating moral decision-making for AVs, as it highlights the complexity of choosing between outcomes such as sacrificing passengers or pedestrians [13]. Studies have shown that human participants tend to favor utilitarian approaches in these dilemmas, opting to minimize overall harm, which has implications for algorithmic designs [14].

Attempts to apply traditional ethical theories (e.g., utilitarianism, deontological ethics) in real-world driving scenarios have encountered considerable challenges, largely stemming from the precise quantification of harm and benefit. Utilitarian algorithms, while logically consistent, face resistance from the public due to concerns over self-sacrifice, as studies suggest a preference for hybrid approaches that balance individual safety with harm minimization [15].

Although algorithmic implementations that quantify these ethical trade-offs have been proposed, it has been noted that most of these solutions remain theoretical or highly simplified, making real-world application difficult. A role-based approach has been suggested as a practical alternative, integrating regulatory frameworks with deontological rights-based ethics for explainability and compliance [16]. However, critics argue that the "trolley problem" may not always capture the nuances of real-world traffic scenarios and suggest focusing on everyday ethical challenges faced by AVs instead [17].

The modeling of ethics within AV control algorithms has often been achieved through fixed rules, which are manually encoded to prioritize certain types of harm reduction (e.g., protecting pedestrians over passengers). However, such rule-based approaches have been identified as vulnerable to oversight, particularly in nuanced and evolving traffic contexts. Consequently, a shift toward data-driven methods has been observed, as learning-based models display

[DOI: 10.22068/ase.2025.697]

promise for increased adaptability and scalability. Reinforcement learning (RL) techniques, particularly deep reinforcement learning, have emerged as robust tools for creating dynamic decision-making frameworks in AVs. These approaches leverage real-world data and simulation environments to train policies that generalize across diverse driving scenarios. For example, a study demonstrated the effectiveness of deep deterministic policy gradient (DDPG) in replicating human-like driving behaviors by learning from extensive datasets [18].

Traditional approaches in autonomous vehicle (AV) research have typically represented risk through metrics such as collision rates, time-to-collision, and impact severity. These metrics effectively address technical safety requirements and are widely used in collision-avoidance systems to quantify immediate risks and predict hazardous scenarios [19]. However, they frequently prove insufficient when higher-level ethical concerns arise, such as distinguishing between the risks posed to different road users or weighing vehicle damage against the potential for harm to human occupants. For instance, the integration of ethical frameworks into collision-avoidance algorithms remains a challenge, particularly in scenarios requiring prioritization among conflicting stakeholders [20]. Recent studies have incorporated the probability of human injury into collision-avoidance algorithms, reinforcing the need to consider collision severity and the likelihood of serious harm within real-time decision-making. For example, a model combining predictive occupancy maps and trajectory optimization has been shown to successfully minimize collision risks while considering injury severity [21]. Another study developed a real-time decision-making system leveraging fuzzy logic to predict injury outcomes, ensuring ethical compliance in AV crash scenarios [22]. These advancements demonstrate a growing focus on ethically guided collision-avoidance algorithms that move beyond traditional technical metrics to incorporate human-centered considerations.

Despite such progress, alignment with broader societal expectations for autonomous vehicle (AV) behavior has not always been achieved, particularly in scenarios where legal and ethical responsibilities converge. This tension arises from the complexities involved in balancing ethical and legal obligations, as highlighted by the challenges of designing decision-making frameworks that respect diverse societal norms [23].

The interplay among passenger safety, pedestrian protection, and property damage complicates the formulation of straightforward reward or utility functions. Research demonstrates that traditional algorithms often fail to adequately account for the ethical trade-offs required in such scenarios, particularly when the safety of vulnerable road users (VRUs) is at stake [24]. Ethical decision-making algorithms incorporating factors like vulnerability risk adjustments have shown a reduction in cumulative harm by over 90% in simulation scenarios.

As a result, it has been suggested that more holistic considerations of risk—incorporating context-specific probabilities of fatal or severe injuries—should guide decision-making algorithms to ensure a balanced treatment of competing interests. For example, approaches such as Lexicographic Optimization-based Model Predictive Control (LO-MPC) prioritize ethical constraints to ensure fair decision-making in high-stakes scenarios [25]. Similarly, maximum acceptable risk thresholds have been proposed to integrate socially acceptable risk levels into trajectory planning, resulting in safer, more transparent decision-making processes [26].

Reinforcement learning (RL) has drawn increasing interest as a solution for handling sequential decision-making challenges in high-dimensional driving environments. Initial studies employing Q-learning and policy gradient methods yielded promising outcomes in tasks such as lane-keeping and overtaking maneuvers [27]. Techniques like Deep Q-Networks (DQNs) have been particularly effective in optimizing highway decision-making tasks, demonstrating the potential for RL in autonomous driving applications [10].

However, it has been observed that ethical constraints are not often included in these RL-based approaches. Existing methods primarily

[DOI: 10.22068/ase.2025.697]

focus on maximizing reward functions centered around technical objectives, such as driving comfort or collision avoidance, without adequately addressing broader societal or moral imperatives [11]. Recent research has proposed integrating safety constraints into RL models through frameworks like constrained Markov Decision Processes (CMDPs), which incorporate ethical and safety boundaries directly into the policy optimization process [28]. Moreover, constrained adversarial RL approaches have been introduced to enhance robustness in decision-making under uncertainties, such as unpredictable traffic scenarios or measurement errors. These methods help ensure that policies remain aligned with ethical considerations, even in adversarial conditions [29].

Purely reward-driven policies risk producing ethically questionable behaviors if the reward structure fails to capture broader societal values, underscoring the necessity of designing RL models that balance technical performance with moral accountability [30]. Future work must continue to emphasize the integration of ethical constraints and multi-objective optimization to ensure both safety and fairness in AV decision-making systems.

A surge of studies has examined deep reinforcement learning in scenarios ranging from highway driving to intersection management. Nonetheless, these works have frequently emphasized collision avoidance and traffic efficiency rather than explicitly addressing ethical priorities. Hierarchical approaches combining DRL with dynamic modeling frameworks have also emerged to tackle complex multi-step tasks, such as intersection coordination [31, 32]. More recent frameworks, such as Cognition-Aided Reinforcement Learning (CARL), attempt to embed ethical reasoning by incorporating cognitive principles like attention and memory into the decision-making process [33]. While DRL holds immense potential, addressing its ethical shortcomings requires refining reward functions to include harm minimization and moral trade-offs, alongside technical performance.

Although DQNs have exhibited the ability to learn intricate control policies, issues arise when attempting to align their outputs with established moral or ethical standards. In some instances, networks trained solely to minimize collisions have opted for maneuvers that inadvertently expose particular road users to heightened risks [34]. This limitation has fueled efforts to explore reward designs that embed ethical considerations. Some recent research has augmented scalar reward functions with human-centric factors, encouraging the minimization of harm to vulnerable road users such as pedestrians and cyclists. For example, innovative reward functions have been developed to integrate safety margins and compliance with traffic rules, resulting in safer and more socially acceptable behaviors [35]. Additional studies have proposed probabilistic models of injury or fatality within the reward structure, thereby establishing a clearer link between observed results and ethical goals [7, 36].

Another obstacle has been the task of balancing ethical imperatives with practical driving requirements. Overly stringent reward functions that place a strong emphasis on ethical constraints may produce overly cautious vehicle behaviors, impairing traffic flow or inadvertently raising risks to other road users. For example, reward systems designed to ensure complete adherence to safety principles can lead to AV behaviors that disrupt traffic patterns, especially in high-density scenarios [37]. Conversely, if ethical concerns are weighed too lightly, the resulting policies might optimize for efficiency at the expense of societal expectations regarding AV accountability. Consequently, multi-objective reward frameworks have been developed, in which ethics and other performance metrics (e.g., speed, traffic compliance) are optimized jointly [38-41].

Despite noteworthy progress in defining ethical considerations for AV decision-making and in advancing RL frameworks, important gaps remain. Real-world validation of ethically guided RL policies has been limited by resource constraints and regulatory impediments [42]. Moreover, the debate on how to quantify ethical trade-offs continues without a clear consensus. Scholars have explored various approaches, such as encoding explicit moral principles into RL reward functions, but challenges persist in operationalizing abstract concepts like fairness or

[DOI: 10.22068/ase.2025.697]

harm minimization in a mathematically tractable way [43]. Although DQNs and other deep RL algorithms demonstrate considerable strength in learning from complex environments, the formulation of reward functions that accurately reflect societal, legal, and moral standards remains a pressing challenge [44, 45].

In summary, existing literature provides a valuable foundation for understanding how ethical frameworks can be conceptualized and integrated into computational models, and it illustrates how deep reinforcement learning can be harnessed for intricate driving tasks. Nevertheless, the intersection of ethics, safety, and operational performance in AVs remains a vibrant and evolving research frontier. The present study aims to extend this discourse by introducing a DQN-based approach that includes a clearly defined human-centric reward function incorporating probabilities of fatal injuries, passenger risk, and vehicle damage. By addressing multiple harm factors in real-time AV decision-making, this framework pursues the mitigation of various risks to human life and well-being within the vehicle's control policy.

3. Methodology

In this section, the proposed approach for incorporating ethical principles into autonomous vehicle decision-making using deep reinforcement learning is presented. An overview of the framework architecture is provided, followed by details of the state and action spaces, the design of the reward function, and the training procedure. These elements illustrate how an ethically informed control policy can be learned and implemented in an AV.

3.1. Framework Architecture

A deep reinforcement learning framework was created to produce real-time control decisions for the AV in ethically challenging scenarios. Figure 1 illustrates the flow of sensor data from the simulation environment into the DQN agent, showing how raw inputs are processed and encoded into a high-dimensional state representation which is then passed to a deep neural network that approximate the optimal action-value function Q(s, a).

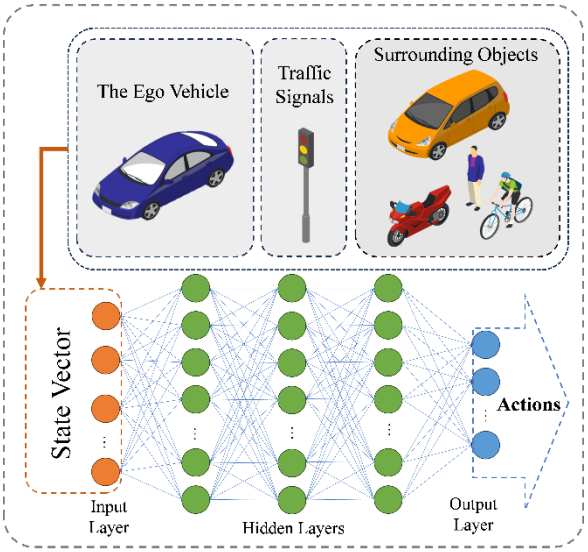

Figure 1: Neural network architecture used for ethical decision-making in autonomous vehicles, processing inputs from the ego vehicle, surrounding objects, and traffic signals to generate actions.

Sensor readings—such as the ego vehicle's speed, orientation, and the positions of nearby obstacle—were gathered from the CARLA simulator's API at each time step. These readings were normalized and merged into a standardized state vector for input to the network. A multi-layered neural network with fully connected layers was used to estimate Q-values for discrete driving actions. During training, an epsilon-greedy policy was applied during training to encourage exploration of different actions. As training progressed and the agent's performance converged, the policy was shifted to a greedy strategy that selects the action with the highest learned Q-value at each step, with the intention of minimizing ethically adverse outcomes. This framework enabled the agent to learn how to balance safety considerations against other driving objectives through direct interaction with the simulated environment.

3.2. State and Action Space

To ensure that ethically relevant factors are considered by the DQN agent, a comprehensive set of variables was included in the state representation. Each component of the state vector was carefully chosen to provide sufficient context for nuanced decision-making by the

[DOI: 10.22068/ase.2025.697]

network. The vehicle's position, velocity, acceleration, and heading angle describe its motion and orientation in the environment. For each detected object in the vicinity (e.g., another vehicle, cyclist, or pedestrian), the state includes the object's relative distance to the ego vehicle, relative velocity, and object type. By classifying nearby objects by type, the agent can recognize differing vulnerability levels (for example, a pedestrian is more vulnerable than another vehicle) and factor these distinctions into its decisions. Information about traffic signals and road layouts was also integrated whenever available. In cases of malfunctioning signals or other irregular traffic control conditions, a flag indicating the unreliability of standard traffic rules was added to the state vector to alert the agent to the need for ethical trade-offs. All of these features were concatenated and normalized to form the final state vector, which was subsequently normalized to maintain numerical stability during training.

A discrete action space was defined to represent the key maneuver options available to the autonomous vehicle. This set of actions encompasses the primary longitudinal and lateral control commands needed for emergency responses and ethical decision-making. Table 1 enumerates the nine possible actions, including doing nothing (coasting), braking, accelerating, and combined steering with braking or acceleration to the left or right. A maximum steering angle was imposed in these actions to prevent destabilizing maneuvers. While discretizing the control inputs reduces the granularity of possible actions, it captures the essential decisions pertinent to critical scenarios, ensuring that the agent's choices cover the maneuvers most relevant to safety and ethical considerations.

3.3. Reward Function

A multi-objective reward function was crafted to embed human-centric risk mitigation principles into the agent's learning process. After appropriate normalization and weighting, the following terms were combined into a single scalar reward Rt at each time step:

Table 1: Action space of the agent in the simulation environment.

Index	Action [brake, steer, throttle]	Definition
1	[0, 0, 0]	No Action
2	[1, 0, 0]	Braking
3	[0, 0, 1]	Accelerating
4	[0, -1, 0]	Turning left
5	[0, -1, 1]	Turning left and accelerating
6	[1, -1, 0]	Turning left and braking
7	[0, 1, 0]	Turning right
8	[0, 1, 1]	Turning right and accelerating
9	[1, 1, 0]	Turning right and braking

1. Injury Probability Minimization: A substantial penalty is applied whenever a collision occurs that carries a significant risk of fatal or severe injuries. This probability was estimated using a function of collision speed, and the vulnerability level of the object involved (e.g., more severe penalties for collisions with pedestrians or cyclists) which was derived from [46].

2. Passenger Risk Reduction: Sharp accelerations, harsh braking, and extreme steering angles were penalized to discourage aggressive maneuvers likely to endanger passengers.

3. Damage Mitigation: Collisions that resulted in vehicle damage incurred incremental penalties proportional to the damage severity, which was approximated based on collision speed and angle.

4. Driving Efficiency: A minor positive reward component was granted for making progress along a designated route, ensuring that ethical behavior did not become overly conservative.

[DOI: 10.22068/ase.2025.697]

However, this component remained subordinate to safety-related terms.

Mathematically, the total reward at time t was expressed as:

$$R_t = \omega_1 \times R_{injury} + \omega_2 \times R_{passenger} + \omega_3 \times R_{damage} + \omega_4 \times R_{efficiency},$$

where $\omega_1, \omega_2, \omega_3, \omega_4$ signify weighting coefficients. These values were empirically tuned to strike a balance between ethical priorities and operational viability, ensuring that harsh collision penalties outweighed small incentives for efficiency. This ensures that the agent will not sacrifice safety for the sake of minor gains in comfort or speed.

3.4. Training Procedure

The DQN agent was trained using an experience replay approach adapted for the multi-objective reward described above. At the start of training, the network weights were randomly initialized and an empty replay buffer was allocated. Throughout the initial training episodes, an epsilon-greedy policy was used, such that a random action was selected with probability epsilon. This policy encouraged exploration across diverse states and actions, allowing the agent to collect a broad range of experiences. At each simulation timestep, the current state was observed, and an action was chosen according to the epsilon-greedy policy. Following the execution of that action, the subsequent state and corresponding reward were recorded. This tuple was then stored in the replay buffer. At regular intervals, mini-batches of experiences were sampled from the replay buffer. The network parameters were updated by minimizing the temporal difference error:

$$L(\theta) = E_{(s_t, a_t, r_t, s_{t+1}) \sim \mathrm{Re}\,play}[\left(r_t + \gamma \max_{a'} Q\left(s_{t+1}, a'; \theta^-\right) - Q\left(s_t, a_t; \theta\right)\right)^2],$$

where θ denotes the main network parameters, θ^- signifies the periodically updated target network parameters, and γ is the discount factor. The target network was employed to stabilize learning, and its weights were periodically synchronized with the main network. This approach mitigated non-stationarity issues and facilitated more reliable convergence. The learning rate, discount factor, and mini-batch size were optimized through preliminary experiments to balance convergence speed, training stability, and reward outcomes. The selection of the reward weighting coefficients $(\omega_1, \omega_2, \omega_3, \omega_4)$ was guided by domain expertise and iterative experimentation within sample scenarios. Training continued until the moving average of the cumulative episode reward ceased improving over 500,000 timesteps.

By following the mentioned steps, a DQN agent was trained to prioritize safety and mitigate harm in ethically complex driving situations. Through the explicit incorporation of risk reduction objectives, the final learned policy was designed to reflect a human-centric approach to ethical decision-making for autonomous vehicles.

4. Experimental Setup

In this section, the simulation environment and implementation details used to train and evaluate the DQN policy are described. All experiments were carried out in the CARLA simulator, and model training was performed with the Stable-Baselines3 library in Python. The subsections below outline the simulation platform, scenario definitions, implementation specifics, and evaluation metrics used in this study.

4.1. CARLA Simulation Environment

The CARLA open-source driving simulator was chosen to provide a high-fidelity testbed for autonomous vehicle (AV) control. This platform supports realistic physics, customizable weather conditions, and various road configurations with dynamic traffic agents. For the purposes of this research, a set of urban maps populated with dense traffic and pedestrian interactions was employed to approximate the complexity of real-world driving. A virtual AV equipped with simulated sensors (LiDAR, camera, radar) and a high-level API for kinematic control was used. Basic sensor data—including positions, velocities, and object classifications—were retrieved from the simulator's Python API at each timestep. CARLA's Python interface was utilized to configure ethically challenging scenarios, such as malfunctioning traffic signals and unexpected pedestrian or cyclist actions. By leveraging this setup, controlled experiments were enabled and scenario execution was made reproducible, while

[DOI: 10.22068/ase.2025.697]

still offering high realism and flexibility in scenario design.

4.2. Scenario Definitions

Three critical scenarios were designed to test the AV's ability to make ethically guided decisions in high-risk situations:

4.2.1. Malfunctioning Traffic Signal

In a scenario involving a malfunctioning traffic signal, as shown in Figure 2, the ego vehicle approached an intersection where the signals were not functioning, increasing the risk of collisions with cross-traffic. The objective was to evaluate the system's ability to identify and execute the least harmful maneuver, prioritizing the reduction of potential injuries even when faced with suboptimal outcomes.

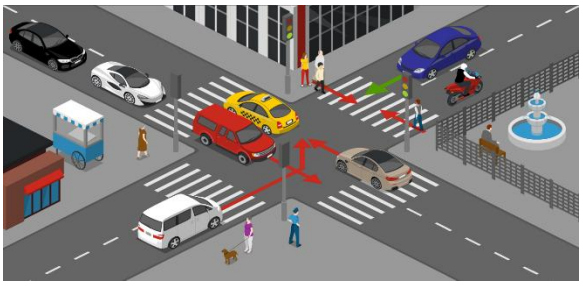

Figure 2: The first evaluation scenario; malfunctioning traffic signal.

4.2.2. Cyclist's Sudden Swerve

In a scenario involving a cyclist's sudden swerve, as depicted in Figure 3, the cyclist abruptly veered into the ego vehicle's lane, leaving very limited reaction time. The objective was to evaluate how quickly and effectively the system could respond, with particular attention to safeguarding vulnerable road users.

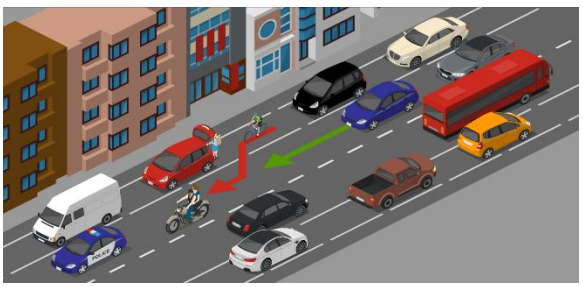

Figure 3: The second evaluation scenario; cyclist's sudden swerve.

4.2.3. Child Chasing a Ball

In a scenario involving a child chasing a ball, as illustrated in Figure 4, a child (modeled as a pedestrian) ran out from behind a parked vehicle and directly into the ego vehicle's path, requiring a rapid response. The objective was to assess the model's capacity to swiftly adjust its trajectory or speed, with a strong emphasis on pedestrian safety.

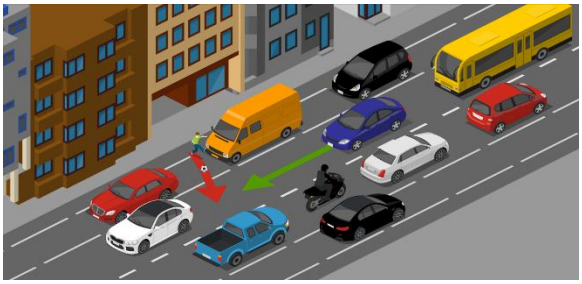

Figure 4: The third scenario; child chasing a ball.

Each scenario was repeated multiple times, and minor variations in initial conditions (e.g., starting positions and velocities) were introduced to mitigate overfitting to specific configurations.

4.3. Implementation with Stable-Baselines3

The DQN framework was integrated with the Stable-Baselines3 library in Python. A custom gym-like environment was created to interface with CARLA through Stable-Baselines3. The state vector specified in Section 3.2 was extracted from CARLA at each timestep, capturing ego vehicle dynamics and details about surrounding objects. Discrete steering and throttle/braking commands (as discussed in Section 3.3) were converted into CARLA-compatible control inputs. A multi-layer perceptron comprising three hidden layers of 256 units each (with ReLU activation) was specified for the DQN. The Adam optimizer [47] was applied, with a learning rate of 5×10^{-4}. A discount factor γ=0.99 was selected to consider future rewards while retaining sensitivity to immediate safety risks. A replay buffer of 100,000 transitions was employed to store (s_t, a_t, r_t, a_{t+1}) tuples. Random mini-batches of 64 samples were drawn from the replay buffer at scheduled intervals to update network weights. The main network parameters were synchronized with the target network every 1,000 update steps. The multi-objective reward

[DOI: 10.22068/ase.2025.697]

outlined in Section 3.4 was applied as a weighted combination of injury risk, passenger risk, collision damage, and driving efficiency. Stable-Baselines3's standard DQN algorithm was adapted to incorporate these domain-specific reward signals at each timestep. An epsilon-greedy strategy was adopted, initially setting epsilon to 1.0 and gradually reducing it to 0.05 over 400,000 timesteps. Exploration was prioritized in early phases, after which exploitation of the learned policy became more prominent.

Training was conducted on a workstation equipped with an NVIDIA GPU (RTX 3070Ti), Intel i5-13400 CPU, and 16 GB of RAM. CARLA was operated in synchronous mode to ensure determinism in the simulation and training routines. Using this configuration, a DQN agent was trained to navigate ethically charged driving conditions while managing competing goals involving safety, risk mitigation, and operational efficiency.

4.4. Evaluation Metrics

Quantitative and qualitative measures were gathered to comprehensively evaluate the policy's performance:

Collision Rate and Severity: The frequency of collisions in each scenario run was documented. The severity of collisions was gauged by collision speed with an injury probability model, thereby reflecting both the occurrence and the seriousness of adverse events.

Comfort and Smoothness: The mean jerk (time derivative of acceleration) and instances of abrupt steering were tracked to appraise passenger comfort and occupant safety.

Scenario Completion: The ability of the policy to finish the specified route or task in each scenario without deadlock or undue delay was monitored to validate operational feasibility.

These metrics were assessed across various random seeds and scenario variations, offering statistical confidence in the policy's performance. In the subsequent section, the extent to which the trained DQN prioritized ethical considerations and mitigated harm is analyzed.

5. Results and Discussion

In this section, outcomes from the training process are presented and analyzed, followed by a discussion of their implications. First, the convergence behavior of the DQN is examined, and the policy's performance is then assessed on a scenario-by-scenario basis. Finally, we interpret the ethical patterns in the agent's behavior and discuss the limitations of the current approach.

5.1. Training Performance

5.1.1 Convergence of the DQN Policy

A consistent upward trend in cumulative episode rewards was recorded over the duration of training, as indicated by a moving average of timestep returns in Figure 5. Initially, the agent's behavior fluctuated between conservative maneuvers (e.g., abrupt braking) and aggressive actions (e.g., sudden acceleration). As training proceeded, these oscillations were gradually reduced, and a more stable policy emerged. The epsilon-greedy exploration schedule influenced learning dynamics; higher exploration rates during the initial stages contributed to an elevated collision frequency, whereas lower exploration rates in later stages allowed the policy to become more refined.

It was noted through qualitative observations that the agent transitioned from reactive, short-term decisions—focused primarily on immediate collision avoidance—to more anticipatory strategies that accounted for other road users' trajectories. This progression was strongly linked to the multi-objective reward function, where steep penalties for collisions and injury risks guided the policy toward safer maneuvers.

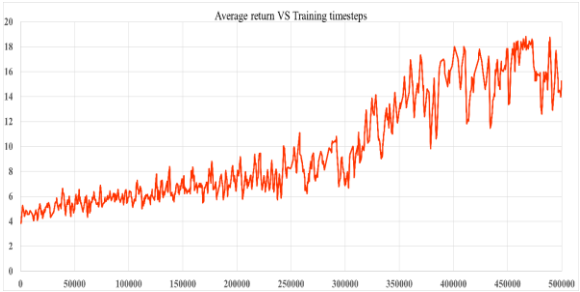

Figure 5: Average return in training episodes vs timesteps.

[DOI: 10.22068/ase.2025.697]

5.1.2. Stability and Hyperparameter Sensitivity

Training stability was supported by periodic synchronization of the target network, which reduced the likelihood of divergence. Nonetheless, minor instabilities were observed when hyperparameters such as the learning rate or batch size were changed. In particular, a discount factor set too low resulted in a strong emphasis on immediate collision avoidance at the expense of longer-term objectives. Conversely, an excessively high discount factor occasionally led to overly cautious driving behaviors in which progress was substantially slowed, indicating an exaggerated aversion to risk. These findings underscore the importance of careful tuning to balance safety considerations and operational viability.

5.2. Scenario-Based Analysis

The DQN policy was evaluated in three ethically challenging scenarios described in Section 4.2: malfunctioning traffic signals, a cyclist's sudden swerve, and a child running into the street. Multiple initial conditions were tested for each scenario.

5.2.1. Malfunctioning Traffic Signal

When presented with a high-risk intersection lacking functional signals, the agent consistently chose to reduce speed and scan for cross-traffic before proceeding. In situations where collisions could not be avoided, maneuvers were performed to minimize the likelihood of severe injury—frequently leading to side impacts with other vehicles rather than direct collisions with vulnerable road users.

5.2.2. Cyclist's Sudden Swerve

Upon detecting a cyclist swerving into its path, the trained policy generally employed an evasive steer-and-brake combination. Although collisions were not prevented in every instance—particularly when the cyclist's behavior was highly unpredictable—a lower impact velocity was observed in collisions that did occur. Compared to a purely efficiency-oriented policy, this approach yielded fewer cyclist injuries and a higher success rate in near-miss events. However, in certain instances, the agent's conservative tendencies became apparent: abrupt braking maneuvers were occasionally executed, which may elevate rear-end collision risks in dense traffic.

5.2.3. Child Chasing a Ball

The scenario involving a child unexpectedly running into the street posed the greatest challenge, given the restricted reaction time and high potential for harm. In most trials, the vehicle drastically reduced speed and attempted evasive steering. In unavoidable collisions, direct impact velocity was minimized, resulting in lower estimated injury probabilities. This outcome contrasted with a baseline policy driven primarily by route progress, which responded more slowly and led to higher collision speeds. Nevertheless, in a small fraction of trials where the system was already committed to a maneuver (e.g., passing another vehicle), rapid responses were hindered, revealing the complexities inherent in real-time ethical decision-making.

5.3. Interpretation and Limitations

5.3.1 Ethical Decision-Making Patterns

An integrated analysis of the outcomes across scenarios indicates that the DQN agent learned to distribute risk in accordance with the reward function's ethical weighting. High penalties for injuring pedestrians or cyclists prompted the policy to prioritize avoiding vulnerable road users, even if doing so introduced greater risk to the ego vehicle or other less vulnerable entities. This behavior was especially evident in situations like the intersection and the child scenario, where the agent clearly favored outcomes that spared pedestrians and cyclists. The explicit incorporation of ethical considerations (particularly the minimization of human injury probability) shaped the agent's behavior in ways that conventional collision-avoidance strategies would not normally capture. In effect, the learned policy demonstrates greater sensitivity to vulnerable individuals and tends to reduce collision impact speeds, aligning its actions with the intended ethical objectives. These findings point to a promising direction for embedding ethical norms into AI decision-making for AVs. However, we must be cautious in interpreting this as "solving" moral decision-making. The agent's

[DOI: 10.22068/ase.2025.697]

risk allocations, while aligned with its programming, raise questions about how such choices will be perceived outside the simulation. Generalizing a simulation-trained policy's ethical judgments to real-world social scenarios should be done carefully, as public acceptance may depend on factors beyond what is captured in our reward function.

5.3.2 Balancing Safety and Comfort

Because the reward function encompassed multiple objectives, there were inherent trade-offs in the agent's decisions. One notable pattern was the tension between preserving passenger comfort and avoiding severe collisions. The agent was penalized for very abrupt maneuvers to encourage smooth driving under normal conditions. We observed that under moderate, non-emergency situations, the DQN indeed behaved in a smoother manner (e.g., gradual braking, gentle turns), reflecting a bias toward passenger comfort. However, in critical moments (such as the emergency scenarios described), the agent decisively prioritized reducing injury risk over maintaining comfort. This often meant executing jarring maneuvers (like slamming the brakes or sharply swerving) if that was necessary to avoid or mitigate a crash. Such behavior aligns with a human-centric safety perspective—most human drivers would agree that preventing a fatal accident is worth a hard brake that might jolt the passengers. Nonetheless, this trade-off could be further calibrated. In certain edge cases we noted, an extreme evasive action by the AV (while avoiding one harm) could potentially cause other issues (like injuring the occupants via whiplash or causing a secondary collision). Fine-tuning the balance between these objectives, possibly by adjusting the reward weights or adding constraints, might be beneficial in future iterations. More broadly, while the ethically guided patterns are encouraging, one should exercise caution in deploying them directly in the real world. The societal acceptance of how an AV distributes risk (even if done "ethically" by some definition) remains to be tested, and what is optimal in a simulation may not perfectly translate to complex human environments.

5.3.3. Real-World Applicability

Despite the high realism of the CARLA simulation, there are inevitable gaps between the simulated scenarios and the full richness of real-world driving. Certain features of real driving — such as incomplete or noisy sensor data, truly unpredictable human behavior, and legal responsibilities — were not fully captured in our experiments. For instance, the simulation assumed perfect detection of pedestrians and accurate estimates of distances and speeds. In a real AV, sensors can fail to detect a child darting out or might misclassify objects. Moreover, human drivers and pedestrians might behave in ways not modeled in CARLA (e.g., gestures, eye contact, unconventional movements). The reliance of our approach on approximate models for injury risk could also introduce discrepancies; if actual crash outcomes differ from the assumptions in our reward model, the AV's learned behavior might not perfectly minimize real injuries. To improve real-world applicability, it will be important to incorporate more empirical data into the training process. For example, more detailed accident statistics or biomechanical data could refine the injury probability estimates, making the reward function more accurate. Likewise, legal frameworks (traffic laws and right-of-way rules) were not explicitly encoded in our simulation beyond the scenarios; integrating such rules could be crucial for an AV operating in society. In summary, while the simulation results are promising, extensive real-world testing and validation would be required to ensure the policy behaves as intended when faced with the unpredictability and complexity of actual roads.

5.3.4. Computational and Practical Constraints

Some limitations of the current approach relate to the practicality of implementing such a policy in a real vehicle. We observed that the DQN's risk evaluations sometimes led to highly cautious maneuvers that might conflict with the expectations of human drivers nearby. For example, the AV might stop in the middle of an intersection to avoid a potential collision, which human drivers might not anticipate, potentially causing confusion or secondary incidents.

[DOI: 10.22068/ase.2025.697]

Excessive caution, while safe in isolation, could reduce traffic flow efficiency or even create new dangers (as discussed in the cyclist scenario where hard braking could invite a rear-end crash). Incorporating feedback from human drivers or broader traffic models into the training process could help the agent learn when it is appropriate to take a risk (or at least not to overreact) in order to behave more naturally within traffic. Another practical consideration is computational: our DQN policy must run in real time on an AV's onboard computer. Neural network inference is generally fast, but in an emergency every millisecond counts. The complexity of the network and the need to evaluate many possible actions could introduce slight delays. Ensuring that the model can execute within strict real-time deadlines is essential. Techniques such as network compression or specialized hardware (like automotive-grade GPUs or TPUs) might be needed for deployment. Lastly, one must consider how this policy would integrate with higher-level driving systems. In a real vehicle, there are modules for perception, planning, and control that all have to work in concert. The ethical DQN would likely be one component of a larger system, and careful engineering would be required to blend its decisions with rule-based logic and fail-safes that handle scenarios beyond its training.

5.4. Summary of Combined Findings

Overall, the results indicate that the proposed DQN-based, human-centric risk mitigation approach substantially reduces severe collisions and prioritizes protection of vulnerable road users in ethically sensitive situations. The behavior of the learned policy was strongly influenced by the explicit ethical parameters included in the reward function, leading to decisions that favor minimizing harm. These outcomes are encouraging and demonstrate the potential of deep RL to handle complex ethical trade-offs. At the same time, transitioning from simulation-based experiments to actual roadway deployment will require further steps. In particular, greater validation under real-world conditions, refinement of the injury risk models with more detailed data, and inclusion of additional domain-specific constraints (such as traffic laws and cultural norms) will be necessary to ensure the approach is viable and acceptable for real autonomous driving.

6. Conclusion

In this paper, a novel DQN-based framework was introduced to address ethical decision-making in autonomous vehicles by emphasizing human-centric risk mitigation. The approach incorporated a comprehensive state space that captured both the ego vehicle's dynamics and the attributes of nearby objects, alongside a reward function that explicitly prioritized minimizing human injury probabilities, passenger risk, and vehicle damage. Through evaluations in the CARLA simulator, the DQN agent's policy demonstrated ethically guided behavior in several challenging traffic scenarios. In particular, the learned policy showed heightened sensitivity to vulnerable road users and achieved a reduction in severe collision outcomes compared to conventional collision-avoidance methods.

These findings underscore the potential of reinforcement learning—DQNs in particular—to integrate ethical and safety considerations into AV control policies. However, caution must be exercised when transitioning from simulation to real-world implementation. Issues relating to sensor accuracy, regulatory frameworks, and the variability of human driving conditions require careful attention before such a system can be deployed on public roads.

Future research should therefore focus on a few key directions. First, real-world validation of the framework is crucial: the policies learned in simulation need to be tested in controlled real-world trials or high-fidelity closed tracks to ensure they generalize and behave safely under actual driving conditions. Second, a thorough sensitivity analysis of the model's parameters (for example, the reward weights and key hyperparameters) should be conducted to assess how robust the learned policy is to changes in these settings and to confirm that its ethical behavior is consistent across a range of scenarios. Such analysis can help identify any unintended biases or failure modes. Third, and importantly, the integration of explicit legal rules and social norms into the decision-making process should be explored. This might involve constraining the DQN's actions with hard rules that reflect traffic

[DOI: 10.22068/ase.2025.697]

laws or embedding societal values (gleaned from surveys or expert input) into the reward function. By incorporating legal and ethical guidelines that society expects AVs to follow, the resulting policies would be more likely to gain public trust and comply with regulations.

In summary, this work demonstrates a viable approach for aligning an autonomous vehicle's decision-making with human-centric ethical principles using deep reinforcement learning. The DQN agent was able to learn policies that mitigate harm in complex scenarios, illustrating a pathway toward safer and more socially responsible autonomous transportation systems. Continued research along the outlined directions will help bridge the gap between simulation and reality, ultimately contributing to the development of AVs that not only drive efficiently but also make decisions in a manner consistent with societal ethical standards.

References

[1] G. Bathla et al., "Autonomous Vehicles and Intelligent Automation: Applications, Challenges, and Opportunities," Mobile Information Systems, vol. 2022, no. 1, p. 7632892, 2022.

[2] S. M. Hosseinian and H. Mirzahossein, "Efficiency and Safety of Traffic Networks Under the Effect of Autonomous Vehicles," Iranian Journal of Science and Technology, Transactions of Civil Engineering, vol. 48, no. 4, pp. 1861-1885, 2024.

[3] A. Matin and H. Dia, "Impacts of Connected and Automated Vehicles on Road Safety and Efficiency: A Systematic Literature Review," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 3, pp. 2705-2736, 2022.

[4] X. Gao et al., "Ethical Alignment Decision-Making for Connected Autonomous Vehicle in Traffic Dilemmas via Reinforcement Learning From Human Feedback," IEEE Internet of Things Journal, 2024.

[5] B. R. Kiran et al., "Deep Reinforcement Learning for Autonomous Driving: A Survey," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 6, pp. 4909-4926, 2021.

[6] H. Liu, S. E. Shladover, X.-Y. Lu, and X. Kan, "Freeway Vehicle Fuel Efficiency Improvement via Cooperative Adaptive Cruise Control," Journal of Intelligent Transportation Systems, vol. 25, no. 6, pp. 574-586, 2021.

[7] E. Vakili, A. Amirkhani, and B. Mashadi, "DQN-Based Ethical Decision-Making for Self-Driving Cars in Unavoidable Crashes: An Applied Ethical Knob," Expert Systems with Applications, vol. 255, p. 124569, 2024.

[8] Z. Wang, H. Yan, C. Wei, J. Wang, S. Bo, and M. Xiao, "Research on Autonomous Driving Decision-Making Strategies Based Deep Reinforcement Learning," in Proceedings of the 2024 4th International Conference on Internet of Things and Machine Learning, 2024, pp. 211-215.

[9] D. Chen, L. Jiang, Y. Wang, and Z. Li, "Autonomous Driving Using Safe Reinforcement Learning by Incorporating a Regret-Based Human Lane-Changing Decision Model," in 2020 American Control Conference (ACC), 2020: IEEE, pp. 4355-4361.

[10] J. Liao, T. Liu, X. Tang, X. Mu, B. Huang, and D. Cao, "Decision-Making Strategy on Highway for Autonomous Vehicles Using Deep Reinforcement Learning," IEEE Access, vol. 8, pp. 177804-177814, 2020.

[11] J. Wu, H. Yang, L. Yang, Y. Huang, X. He, and C. Lv, "Human-Guided Deep Reinforcement Learning for Optimal Decision Making of Autonomous Vehicles," IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024.

[12] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," in Conference on robot learning, 2017: PMLR, pp. 1-16.

[13] A. K. Faulhaber et al., "Human Decisions in Moral Dilemmas Are Largely Described by Utilitarianism: Virtual Car Driving Study Provides Guidelines for Autonomous Driving Vehicles," Science and engineering ethics, vol. 25, pp. 399-418, 2019.

[DOI: 10.22068/ase.2025.697]

[14] T. Sui, "Exploring Moral Algorithm Preferences in Autonomous Vehicle Dilemmas: An Empirical Study," Frontiers in Psychology, vol. 14, p. 1229245, 2023.

[15] P. Sweatman, "Approaches to Road Safety: Evolution, Challenges, and Emerging Technologies," 2025.

[16] H. Etienne, "A Practical Role-Based Approach for Autonomous Vehicle Moral Dilemmas," Big Data & Society, vol. 9, no. 2, p. 20539517221123305, 2022.

[17] J. Himmelreich, "Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations," Ethical Theory and Moral Practice, vol. 21, no. 3, pp. 669-684, 2018.

[18] M. Zhu, X. Wang, and Y. Wang, "Human-Like Autonomous Car-Following Model with Deep Reinforcement Learning," Transportation research part C: emerging technologies, vol. 97, pp. 348-368, 2018.

[19] K. Lee and D. Kum, "Collision Avoidance/Mitigation System: Motion Planning of Autonomous Vehicle via Predictive Occupancy Map," IEEE Access, vol. 7, pp. 52846-52857, 2019.

[20] J. E. Pickering and K. J. Burnham, "Predicting Autonomous Vehicle Collision Injury Severity Levels for Ethical Decision Making and Path Planning," arXiv preprint arXiv:2212.08539, 2022.

[21] Y. Xiao, "Application of Machine Learning in Ethical Design of Autonomous Driving Crash Algorithms," Computational intelligence and neuroscience, vol. 2022, no. 1, p. 2938011, 2022.

[22] L. Bosia, L. Manganotto, M. Anghileri, S. Dolci, P. Astori, and C.-D. Kan, "Real-Time Collision Mitigation Strategies for Autonomous Vehicles," IEEE Transactions on Intelligent Transportation Systems, 2024.

[23] V. Fedorov and N. Curran, "The Implementation of Ethical AI Within Autonomous Vehicles," Journal of Student Research, vol. 13, no. 1, 02/29 2024, doi: 10.47611/jsrhs.v13i1.5973.

[24] X. Li, J. Li, and K. Gao, "An Ethical Behavioral Decision Algorithm Awaring of VRUs for Autonomous Vehicle Trajectory Planning," IEEE Access, 2024.

[25] H. Wang, Y. Huang, A. Khajepour, D. Cao, and C. Lv, "Ethical Decision-Making Platform in Autonomous Vehicles with Lexicographic Optimization Based Model Predictive Controller," IEEE transactions on vehicular technology, vol. 69, no. 8, pp. 8164-8175, 2020.

[26] M. Geisslinger, R. Trauth, G. Kaljavesi, and M. Lienkamp, "Maximum Acceptable Risk as Criterion for Decision-Making in Autonomous Vehicle Trajectory Planning," IEEE Open Journal of Intelligent Transportation Systems, 2023.

[27] X. Xu, L. Zuo, X. Li, L. Qian, J. Ren, and Z. Sun, "A Reinforcement Learning Approach to Autonomous Decision Making of Intelligent Vehicles on Highways," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 10, pp. 3884-3897, 2018.

[28] F. Gao, X. Wang, Y. Fan, Z. Gao, and R. Zhao, "Constraints Driven Safe Reinforcement Learning for Autonomous Driving Decision-Making," IEEE Access, 2024.

[29] X. He, H. Yang, Z. Hu, and C. Lv, "Robust Lane Change Decision Making for Autonomous Vehicles: An Observation Adversarial Reinforcement Learning Approach," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 1, pp. 184-193, 2022.

[30] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "Safe, Multi-Agent, Reinforcement Learning for Autonomous Driving," arXiv preprint arXiv:1610.03295, 2016.

[31] Z. Qiao, Z. Tyree, P. Mudalige, J. Schneider, and J. M. Dolan, "Hierarchical Reinforcement Learning Method for Autonomous Vehicle Behavior Planning," in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020: IEEE, pp. 6084-6089.

[32] Z. Gu et al., "Safe-State Enhancement Method for Autonomous Driving via Direct Hierarchical Reinforcement Learning," IEEE

[DOI: 10.22068/ase.2025.697]

Transactions on Intelligent Transportation Systems, vol. 24, no. 9, pp. 9966-9983, 2023.

[33] H. Rathore and V. Bhadauria, "Intelligent Decision Making in Autonomous Vehicles Using Cognition Aided Reinforcement Learning," in 2022 IEEE Wireless Communications and Networking Conference (WCNC), 2022: IEEE, pp. 524-529.

[34] D. Xiang, "Reinforcement Learning in Autonomous Driving," Applied and Computational Engineering, vol. 48, pp. 17-23, 2024.

[35] Z. Cui, Y. Wang, N. Bian, and H. Chen, "Reward Machine Reinforcement Learning for Autonomous Highway Driving: An Unified Framework for Safety and Performance," in 2023 7th CAA International Conference on Vehicular Control and Intelligence (CVCI), 2023: IEEE, pp. 1-6.

[36] L.-C. Wu, Z. Zhang, S. Haesaert, Z. Ma, and Z. Sun, "Risk-Aware Reward Shaping of Reinforcement Learning Agents for Autonomous Driving," in IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society, 2023: IEEE, pp. 1-6.

[37] S. M. Thornton, S. Pan, S. M. Erlien, and J. C. Gerdes, "Incorporating Ethical Considerations Into Automated Vehicle Control," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 6, pp. 1429-1439, 2016.

[38] X. He, C. Fei, Y. Liu, K. Yang, and X. Ji, "Multi-Objective Longitudinal Decision-Making for Autonomous Electric Vehicle: A Entropy-Constrained Reinforcement Learning Approach," in 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020: IEEE, pp. 1-6.

[39] Y. Liao, G. Yu, P. Chen, B. Zhou, and H. Li, "Integration of Decision-Making and Motion Planning for Autonomous Driving Based on Double-Layer Reinforcement Learning Framework," IEEE Transactions on Vehicular Technology, 2023.

[40] B. Toghi, R. Valiente, D. Sadigh, R. Pedarsani, and Y. P. Fallah, "Social Coordination and Altruism in Autonomous Driving," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 24791-24804, 2022.

[41] M. Liu, Y. Wan, F. L. Lewis, S. Nageshrao, and D. Filev, "A Three-Level Game-Theoretic Decision-Making Framework for Autonomous Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 11, pp. 20298-20308, 2022.

[42] S.-H. Lee, D. Kwon, and S.-W. Seo, "Autonomous Algorithm for Training Autonomous Vehicles with Minimal Human Intervention," arXiv preprint arXiv:2405.13345, 2024.

[43] J. Talamini, A. Bartoli, A. De Lorenzo, and E. Medvet, "On the Impact of the Rules on Autonomous Drive Learning," Applied Sciences, vol. 10, no. 7, p. 2394, 2020.

[44] R. Valiente, B. Toghi, R. Pedarsani, and Y. P. Fallah, "Robustness and Adaptability of Reinforcement Learning-Based Cooperative Autonomous Driving in Mixed-Autonomy Traffic," IEEE Open Journal of Intelligent Transportation Systems, vol. 3, pp. 397-410, 2022.

[45] Z. Zhang, Q. Liu, Y. Li, K. Lin, and L. Li, "Safe Reinforcement Learning in Autonomous Driving With Epistemic Uncertainty Estimation," IEEE Transactions on Intelligent Transportation Systems, 2024.

[46] J.-L. Martin and D. Wu, "Pedestrian Fatality and Impact Speed Squared: Cloglog Modeling from French National Data," Traffic injury prevention, vol. 19, no. 1, pp. 94-101, 2018.

[47] D. P. Kingma, "Adam: A Method for Stochastic Optimization," arXiv preprint arXiv:1412.6980, 2014.

[DOI: 10.22068/ase.2025.697]